

The role of auditory beats induced by frequency modulation and polyperiodicity in the perception of spectrally embedded complex target sounds

Cécile M. H. Marin

Institut de Recherche et de Coordination Acoustique/Musique (IRCAM), 1 place Igor-Stravinsky, F-75004 Paris, France

Stephen McAdams

Laboratoire de Psychologie Expérimentale (CNRS), Université René Descartes, EPHE, 28 rue Serpente, F-75004 Paris, France, and IRCAM, 1 place Igor-Stravinsky, F-75004 Paris, France

(Received 17 January 1995; revised 4 April 1996; accepted 10 April 1996)

The contribution of auditory beats to the perception of target sounds differing from an interfering background by their frequency modulation (FM) pattern or by a difference in fundamental frequency (F_0) was investigated. On each trial, test sounds composed of a single, second-order formant were embedded in harmonic backgrounds and presented in successive intervals. The center frequencies of these “normal” formants differed across intervals. Subjects were to decide which interval contained the test formant with a center frequency matching that of an isolated target formant presented before each test stimulus. Matching thresholds were measured in terms of the width of modulation for FM stimuli or the mistuning of the F_0 's of unmodulated test formants relative to that of the background. Beats may have allowed the identification of the spectral region of the target in both experiments. To reduce interactions between test and background components, matching thresholds were measured for “flat” formants composed of two or three equal-amplitude components embedded in a harmonic background in which components corresponding to those of test formants were absent. These measures were repeated with the addition of a pink noise floor. Matching was still possible in all cases, though at higher thresholds than for normal formants. Computer simulations suggested that the modulation depth of envelope fluctuations within auditory channels played a significant role in the matching of target sounds when their components were mixed in the same frequency region with those of an interfering sound, but not when the target and background components were separated by as much as 250 Hz, the F_0 of the stimulus. © 1996 Acoustical Society of America.

PACS numbers: 43.66.Mk, 43.66.Jh, 43.66.Lj, 43.66.Dc [LLF]

INTRODUCTION

Two frequency-based stimulus parameters have been studied as potential contributors to the perceptual segregation of concurrent harmonic sound sources, namely the coherence of frequency modulation patterns in the two sounds and the difference in their fundamental frequencies. Helmholtz (1885) advanced the idea that nonparallel pitch movements help the auditory system to separate concurrent sound sources. Coherent modulation maintains the frequency ratios among components. It has been shown that the presence of frequency modulation (FM) contributes to the segregation of several concurrent harmonic or inharmonic spectra, whether they are all modulated coherently or not (Bregman and Doehring, 1984; Marin and McAdams, 1990; McAdams, 1984, 1989). However, recent work demonstrates that: (1) subjects are not able to use FM coherence to group frequency components coded in separate auditory channels in the absence of some form of within-channel cue (McAdams and Marin, 1990); (2) they cannot detect whether FM across auditory channels is coherent or incoherent in the absence of a within-channel cue (Carlyon, 1991, 1992); and (3) in the segregation of simultaneously presented vowels, the contribution of FM may be due primarily either to the inharmonic-

ity or polyperiodicity that it produces (Summerfield and Culling, 1992) or to an increased salience of modulated vowel components (Culling and Summerfield, 1995).

Polyperiodicity (Marin, 1991), or the presence of several periodic sounds with different periods that are not integer multiples of one another, has also been tested by several investigators. Polyperiodicity is usually produced experimentally by introducing a difference between the fundamental frequencies (F_0) of two or more harmonic complexes. The identification of two simultaneous vowels improves as the difference between their respective F_0 's (ΔF_0) increases from zero to one or two semitones (6%–12%) (e.g., Schefers, 1983; see de Cheveigné *et al.*, 1995 for a review). However, Chalikia and Bregman (1989) found that segregation was more difficult for a ΔF_0 of an octave than for a separation of 0.5 or 6 semitones. This result suggests that it is not simply ΔF_0 , but perhaps polyperiodicity or harmonic non-coincidence (de Cheveigné, 1993) that is responsible for the phenomenon: The combination of two harmonic sounds separated by an octave gives a waveform with a single period. Indeed, work by Marin (1996) is consistent with this hypothesis.

The research reported here studied the possibility that

ΔF_0 and FM may give rise to other auditory cues that can be used in situations in which listeners are trying to decide whether or not a target sound with known properties is embedded in a complex background sound. Two of the experiments reported below (experiments 1 and 4) formed part of a larger series of experiments that employed a task in which subjects were to match one of two test formants embedded within background sounds to an isolated target formant (Marin, 1991). These two experiments were concerned with the segregation cues frequency modulation and polyperiodicity, respectively. FM may reduce to polyperiodicity in stimuli with few components in which those with different FM patterns do not stimulate the same auditory filters. Further, polyperiodicity is a potent cue for perceptual segregation of concurrent harmonic sounds. However, both cues may also provide important temporal information that can be used by top-down selection processes in deciding whether or not a given target sound is likely to be embedded in a given background sound even in cases where that sound is not fully segregated. Indeed, while conducting experiments 1 and 4 of the present series, it was noted that for both cues, the components of the test formant and background sounds interacted with one another and created auditory beats, or roughness. The research presented here investigated the contribution to target matching performance of amplitude envelope fluctuations produced by these parameters. Accordingly, four additional experiments were performed. Experiments 2 and 3 studied frequency modulation in stimuli in which beats were considerably reduced. Experiment 5 investigated the polyperiodicity cue with stimulus conditions similar to those in experiment 3. In addition, computer simulations of peripheral auditory filtering were performed to model the degree to which auditory beating was present in the different stimulus conditions investigated. Finally, experiment 6 sought to determine the amplitude modulation depth at which subjects were able to detect the frequency region in which the interactions occurred when AM was the only available cue. These latter results are then used to evaluate the potential contribution of envelope fluctuations in the five other experiments.

I. GENERAL METHOD

Each trial consisted of two pairs of sounds presented sequentially. The first sound of each pair was a target formant presented in isolation and the second sound contained a test formant with either the same or a different formant center frequency and was embedded in a harmonic background. The subject was required to designate the pair in which the isolated target and embedded test formants were identical. Frequency modulation depth and mistuning of the test formant F_0 relative to that of the background were varied and threshold target matching performance was determined using an adaptive tracking procedure.

A. Stimuli

The stimuli for the experiments presented here were composed of a formant and a sound whose function was to be a background. In order to obtain a stimulus in which the spectral envelope would give no information about the em-

bedded formant in the absence of FM or ΔF_0 cues, the backgrounds were designed to have spectral envelopes that were the inverse of those for their respective test formants. Both test formants and background sounds were derived from a harmonic series with a nominal F_0 of 250 Hz.

1. Single-formant test sounds

Single-formant sounds were used whose central frequencies (CF) were situated within the range of frequencies found for the first three formants of the vowels /i/, /e/, /a/, /o/, /u/. The spectral envelopes for the formants in experiments 1 and 4 were derived from those of second-order bandpass filters. The CFs used were 325, 700, 1150, and 1700 Hz. Their respective bandwidths were 50, 60, 90, and 100 Hz measured at -3 dB from peak or 130, 160, 240, and 265 Hz measured at -10 dB from peak.

The formants used in experiments 2, 3, 5, and 6 were "flat" formants composed of two or three equal-amplitude harmonics. Their center frequencies were 375, 750, 1250, and 1750 Hz. These center frequencies were chosen to be close to those of the "normal" formants. The 375-Hz formant was the only one to have two components and its CF was the mean of the frequencies of the first two harmonics.

2. Background sounds

Each test sound had a different background. For the normal formants in experiments 1 and 4, the formant peak in the test sound was a valley in the background and vice versa. In this way, if the formant and the background had the same F_0 and starting phase and neither was modulated in frequency, the global spectral envelope would be flat.

The backgrounds for the "flat" formants were designed in the same way. They were composed of the harmonic series with components situated on each side of the components of the flat formant, but were completely missing the components corresponding to those of the formant.

3. General characteristics

The duration of each sound was 1 s including 200-ms linear rise and fall ramps. The rise and fall of test formant and background were synchronous. The combined spectrum contained 16 harmonics added in sine phase, the highest frequency thus being 4 kHz.

With the exception of stimuli for experiment 6, sound files were synthesized with the Csound program (Vercoe, 1986) on a VAX 11/780. An additive synthesis algorithm was used in which each frequency component was generated by a digital oscillator whose amplitude and frequency were controlled in a continuous fashion. The spectral envelopes for the stimuli of experiments 1 and 4 were stored as interpolated table look-up functions, the amplitude of each component being determined by its instantaneous frequency. The stimulus waveforms were digitally synthesized at a sampling rate of 10 kHz. All calculations took place in 32-bit floating-point format and the waveform was then stored in 16-bit integer format on disk. Each stimulus was transferred to the hard disk of a Macintosh II which controlled the experiment.

For experiment 6, stimuli were synthesized in real-time at a sampling rate of 12 kHz with a DSP card (Smith and Chervin, 1986) controlled by the Macintosh II. Calculations took place in 24-bit integer format and were output in 16-bit integer format.

In all experiments, digital waveforms were converted to analog signals through Burr–Brown 706 DACs. The signal was passed through two Rockland 432 low-pass filters in series, each giving -48 -dB/oct attenuation. The cutoff frequencies were set at 40% of the sampling rate. The filtered signal was amplified with an MB Systems 105a power amplifier and presented diotically over Beyer DT48 earphones at a level of 75 dBA (experiments 1, 2, 4, and 6) or 59 dBA (experiments 3 and 5). To verify the presentation level, each earphone was connected via a flat-plate coupler to a Bruel & Kjaer 1209 sound level meter (A-weighting). The subject was seated in a Soluna SN1 double-walled sound attenuation booth during the experiment.

B. Experimental procedure

Two pairs of tones were presented in each trial separated by a 400-ms silence. Each pair was composed of an unmodulated, isolated formant that we call the target formant, followed by a 200-ms silence and then a compound stimulus comprising a test formant and its corresponding background. The test formant differed from the background in a way that depended on the experiment. In experiments 1–3, its components were modulated sinusoidally in frequency while those of the background were unmodulated. In experiments 4 and 5, it differed statically in F_0 from the background. In experiment 6, its components were modulated in amplitude while those of the background were unmodulated. In one pair, the target and test formants had the same CF, and in the other, the test formant was the next highest neighbor of the target formant in terms of CF. In experiments 1 and 2 for example, if the target CF was 700 Hz, the next highest neighbor's CF was 1150 Hz. The isolated targets presented in the two pairs were identical. The order of the two pairs was randomized. Subjects were required to decide if the pair in which the target and test formants were the same was in the first or second interval (2I,2AFC) and to press the appropriate button on a response box. They were told that their judgments should focus on detecting the timbre of the target formant embedded in the background. Feedback was given concerning the correct response by way of lights on the response box.

A one-up/two-down tracking procedure (Levitt, 1971) was used to determine the 70.7% threshold of target matching. The dependent variable (percent rms frequency modulation width, difference in fundamental frequency, or rms amplitude modulation index) was varied linearly as a function of subjects' responses. For each target formant, 12 or 16 turnarounds in the adaptive procedure were recorded. The first four were discarded, and the means and standard deviations of the remaining 8 or 12 were calculated. The mean was taken as an estimate of the matching threshold for the run.

During early testing, it became apparent that different subjects did not have the same thresholds. In order to avoid

experimental runs that were too long, due to a small step size or a large range for the adaptive procedure, these parameters were fixed separately for each subject in each experiment and are reported below. The tracking procedure started at the specified maximum value and remained there until the subject made a correct response. When a subject's tracking stayed too close to the top or bottom of the range, the run was rejected and the values were readjusted.

Within each experimental run, the trials composing the tracking procedure for the different target formants were interleaved in a random fashion. If threshold was determined for a given target before that of the others, trials containing that target at the threshold value continued to be presented, though further data were not recorded for it.

At the beginning of each run, subjects were allowed to listen at will to pairs of stimuli that were to be presented in the experiment, so that they could easily associate the isolated target with the test formant of identical CF embedded in the background and distinguish it from the embedded test formant with a higher CF. The target CF could be changed by pressing a button. These pairs were presented at the highest value of the dependent variable that had been fixed for that subject.

A strong learning effect was observed at the outset of testing. Therefore subjects were trained for several runs until their thresholds appeared to stabilize. Only the last two (experiments 1–3) or five (experiments 4–6) runs for each experiment were retained for further analysis.

C. Subjects

Eight subjects with normal hearing according to self-report (including the two authors) participated in experiments 1–3 and were paid for their services on an hourly basis (excluding the authors). Five of these (including the authors) participated in experiments 4 and 5. Only the authors served as subjects in experiment 6. The ages of the subjects ranged from 18 to 36 years (mean 24.7 years). Five were female and three were male.

II. EXPERIMENTS ON FREQUENCY MODULATION

A. Experiment 1: Normal formants

1. Stimuli and method

The components of the embedded formant were modulated sinusoidally at a rate of 5 Hz. The modulation waveform always started in sine phase. Target formant CFs of 325, 700, and 1150 Hz were used. These could be compared with test formant CFs of 325, 700, 1150, and 1700 Hz. The percent rms modulation width required for threshold was determined from the last 12 of 16 turnarounds in the adaptive procedure for eight subjects.

Starting (maximum) value and stepsize for the adaptive procedure were adjusted for each subject during the training phase. The range of maximum values in the final two runs retained for analysis was 1.06%–4.24%. The range of stepsizes was 0.07%–0.28%. For two subjects different maxima and stepsizes were used for the two lowest and for the highest target formant.

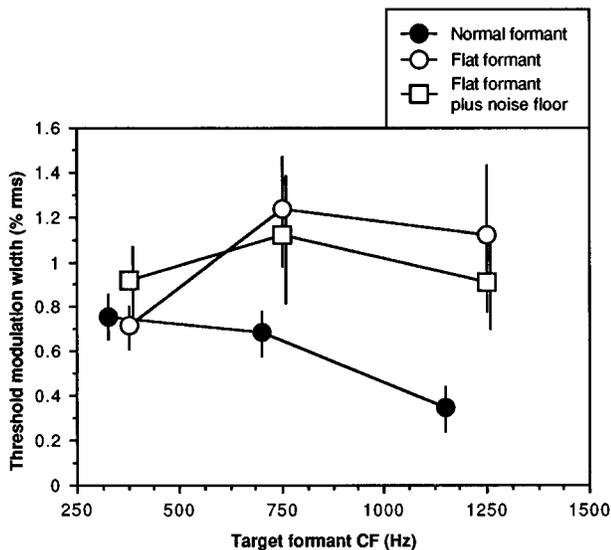


FIG. 1. Summary of frequency modulation results for experiments 1–3 (normal formants, flat formants, and flat formants plus noise floor, respectively). Mean target matching thresholds, expressed as percent rms modulation width, are plotted as a function of target formant CF. Some means do not include values for subjects unable to attain threshold (see text). Vertical bars represent \pm one standard error of the mean.

2. Results

The means of the last two runs across subjects are presented in Fig. 1 (filled circles) for each formant. Matching thresholds for all targets were situated between 0.13% and 1.46% rms frequency modulation width. Formants with lower CFs had higher thresholds (0.76% and 0.69%) than the one with the highest CF (0.35%) as revealed by Tukey–Kramer comparisons on means for each formant across subjects and repetitions (the critical difference at $p=0.05$ was 0.32).

3. Discussion

The thresholds in this experiment were higher than the mean frequency modulation detection thresholds (FMT) found in studies that did not have embedded components: 0.09% rms by Hartmann and Klein (1981) and 0.18% rms by McAdams (1984, Appendix D). Some thresholds, however, were lower than the mean 0.44% rms threshold measured by Demany and Semal (1989). Thus, with one possible exception, the FM should have been easily detectable at matching threshold, though detectability is clearly not sufficient to perform the present task.

The matching thresholds found here are roughly equivalent to detection thresholds found for a frequency modulation incoherence discrimination task by McAdams and Marin (1990) for similar multicomponent targets. The task in that experiment was to detect which of two intervals contained incoherent, aperiodic modulation functions on two subsets of components. In the earlier study however, thresholds increased for higher CF targets rather than decreasing as in the present experiment. The thresholds of the current experiment are smaller by a factor of 3 than Carlyon's (1991, experiment 4) thresholds for discrimination of sinusoidal FM incoher-

ence due to a modulation phase delay on the middle component of a three-component complex.

It is likely that the perception of low-rate beating induced by frequency modulation in this experiment was a contributing factor in target formant matching by allowing subjects to detect the frequency region where the formant was situated. Low-rate beating has several sources in these stimuli, three of which include the following.

- (1) Within-channel interactions between nearby test and background components, the beat frequency of which would vary over time as the components converged, crossed, and diverged again. For a given difference between the fundamental frequencies of test formant and background, the beat rate would be higher for higher CF formants since the maximal frequency difference is proportional to the harmonic rank.
- (2) Amplitude modulation induced by tracing the formant spectral envelope, which would have the greatest modulation depth on the flanks of the formant peak.
- (3) Amplitude modulation induced by components situated in the skirts of auditory filters.

The first two of these sources can be reduced by replacing the second-order “normal” formants by two- or three-component “flat” formants. The frequency separation between test and background components is thus increased and component amplitudes no longer change with F_0 (i.e., envelope tracing is eliminated). This modification was studied in Experiment 2.

B. Experiment 2: “Flat” formants

1. Stimuli and method

Experiment 1 was replicated with three “flat” formants as targets. All eight subjects in this experiment had participated in experiment 1. The starting (maximum) values of rms FM width used in the adaptive procedure ranged from 1.0% to 4.2%. The stepsizes ranged from 0.14% to 0.28%.

2. Results

Matching thresholds are presented in Fig. 1 (open circles). Even when low-rate beats were greatly reduced, subjects were still able to match the correct embedded test formant to the target formant. Individual mean thresholds varied between 0.19% and 3.43%. The effect of target formant CF varied a great deal across subjects. The highest formant had the lowest threshold for three subjects, but had the highest threshold for three others. A threshold was not measurable for one subject with the 1250-Hz target and for another with the 375-Hz target, even after several hours of training and maximum rms FM widths of up to 5.7%. In Tukey–Kramer comparisons between thresholds for the three formants, none of the global differences were found to be significant (the critical difference at $p=0.05$ was 0.76 for 375 vs 1250 CFs and 0.73 for the two other comparisons).

In order to compare the data from experiments 1 and 2 (Fig. 1, filled versus open circles), an analysis of variance with factors formant type (normal versus flat formants), target formant CF, and repetitions were performed on the mean

matching thresholds for all subjects (with missing data for two subjects in experiment 2). The main effect of formant type was highly significant [$F(1,80)=9.95$, $p<0.005$] and the interaction between formant type and target formant CF just attained significance [$F(2,80)=3.14$, $p<0.05$]. Planned comparisons between experiments for each formant CF showed no difference between the 325-Hz formant and the 350-Hz flat formant [$F(1,80)=0.029$, n.s.], a significant difference between the 700-Hz formant and the 750-Hz flat formant [$F(1,80)=5.77$, $p<0.05$], and a highly significant difference between the 1150-Hz formant and the 1250-Hz flat formant [$F(1,80)=10.58$, $p<0.005$].

3. Discussion

The difference in formant shape had an effect on thresholds only for the highest CFs. This effect can be explained in two ways. The larger spacing between formant and background components in flat-formant stimuli eliminated the low-rate beats between them. Auditory channels in the vicinity of higher CFs contain a larger number of stimulus components, so it is logical that thresholds for those CFs were more affected by the change in formant shape. The amplitude of AM induced by envelope tracing in experiment 1 was also largest at high CFs, so the elimination of that effect should mostly have affected thresholds of high-CF formants. These results thus suggest that FM-induced AM cues produced by either of the two mechanisms, or both, were involved in formant matching. The fact that matching was still possible (albeit reduced) for flat formants suggests that the third mechanism of FM-induced AM may also be involved.

Interactions between adjacent components of the modulated test formant and unmodulated background that stimulated the same auditory channels may have persisted in the stimuli of the present experiment, particularly for the higher CF formants. They would have a relatively high rate around 250 Hz with a superimposed 5-Hz periodicity due to the varying mistuning of the harmonic ratios by the FM. In the auditory channels in the region of the modulated test components, a further cue could be fluctuations in activity induced by the back and forth motion of the excitation envelope. The lack of change of thresholds for the lowest CF target suggests that envelope tracing and local component beating did not contribute in the first place at this CF or that another, equally potent, cue, such as AM induced by the auditory filters, was used. The contribution of this source of AM can be evaluated by comparing the FM data with those obtained using a static mistuning (see Sec. III).

In the stimuli of the present experiment, beats may also have been caused by the presence of combination tones (CT) (Goldstein, 1967; Plomp, 1976): nonlinear distortion products created by the background components may have interacted with the test formant components and vice versa. One can eliminate the difference tone ($f_2 - f_1$) by presenting the stimuli at a spectrum level below 50 dB SPL. The level of the most intense higher-order CT ($2f_1 - f_2$), when f_1 and f_2 are adjacent components, is less than that of either component by about 15 dB and could thus be masked by a noise floor (Plomp, 1976). Further, given the level difference between stimulus components and possible CTs, any beating

produced would be very weak. The possible further reduction of beats between partially resolved test and background components and between stimulus components and CTs by the introduction of a noise floor was investigated in experiment 3.

C. Experiment 3: “Flat” formants in the presence of a pink noise masker

1. Stimuli and method

Experiment 2 was replicated with a level of 35 dB/component. A pink noise floor (slope of -3 dB/oct in the power spectrum) was added to the harmonic signal. Its spectrum level was estimated at approximately 8 dB at 250 Hz, 5 dB at 500 Hz, 3 dB at 1 kHz, and 0 dB at 2 kHz. The estimated level was 25.1 dB in an equivalent rectangular band centered on 250 Hz, 23.5 dB on 500 Hz, 23.7 dB on 1 kHz and 23.8 dB on 2 kHz. If the level of cubic difference tones in the vicinity of the harmonic components was 20 dB, the majority of them would be just at masked threshold, presuming about a 4-dB surplus in noise power is necessary for masking. The noise and harmonic signals were played simultaneously and were gated synchronously with 200-ms linear ramps. All eight subjects had participated in experiments 1 and 2. The starting (maximum) values of rms FM width used in the adaptive procedure ranged from 1.1% to 4.2%. The stepsizes ranged from 0.14% to 0.28%.

2. Results

The data are summarized in Fig. 1 (open squares). Individual mean thresholds for target formant matching varied between 0.22% and 3.29%. A threshold could not be measured for one subject with the 375-Hz target. Tukey–Kramer comparisons among thresholds for the different target CFs revealed no significant differences (the critical difference at $p=0.05$ was 0.94 for 750 vs 1250-Hz CFs and 0.97 for the two other comparisons).

A comparison of the data for experiments 2 and 3 (Fig. 1, open circles and squares) revealed that the introduction of a noise floor did not increase thresholds. An analysis of variance (with missing data for two subjects from experiment 2 and one from experiment 3) on factors stimulus type (with or without noise floor), target formant (3 CFs), and repetitions (2) confirmed that neither the main effect of stimulus type nor its interaction with formant CF were significant [$F(1,78)=0.07$, $F(2,78)=0.52$, respectively].

3. Discussion

The inclusion of a noise floor to mask combination tones and small regions of interaction between adjacent, partially resolved components did not produce an increase in thresholds for flat formant stimuli. There are two possible reasons for this result. One is that the matching of target formants measured in the previous experiment does not result from these cues. The other is that the noise floor did not reduce low-rate beating cues due to CTs or higher-rate beating cues resulting from time-varying mistuning of harmonic relations between test and background components. The presentation

of noise at a level sufficient to mask the CTs may not guarantee that they no longer created beating interactions with the audible stimulus components.

The possibility that cues such as AM induced by auditory filter skirts resulting from frequency modulating the test formant components contributed to target matching can be partially evaluated by comparing these data with those obtained from similar stimuli in which a static mistuning is applied to the components of the test formant. This stimulus modification was performed in experiments 4 and 5.

III. EXPERIMENTS ON POLYPERIODICITY

A. Experiment 4: Normal formants

1. Stimuli and method

The F_0 of the embedded test formant was equal to or greater than that of the background. The isolated target formant was not shifted in F_0 . Five subjects participated in the experiment, all of whom had taken part in experiment 1.

Certain modifications were made to shorten the procedure in this experiment. Only 12 turnarounds were recorded in the adaptive procedure, the last eight being used to compute the threshold. The step size in the first two turnarounds was four times the final one and that in the second two turnarounds was two times the final one. The final stepsize remained constant for the last eight turnarounds. In addition, only the isolated target CFs of 325 and 1150 Hz were studied. Subjects began by doing six runs. The results of the first run were not analyzed. If the interval between the highest and lowest thresholds over five consecutive runs was greater than a criterion value of 1.5 times the final stepsize, the subject was required to complete additional runs until this criterion was met. The analysis was performed on the last five runs.

The starting (maximum) values of ΔF_0 used in the adaptive procedure were 0.2% and 2.0%. The stepsizes were 0.01% and 0.1%. For four subjects, the lower values were used for the 1150-Hz formant and the higher values for the 325-Hz formant. For one subject the lower value was used for both formants.

2. Results

Mean threshold F_0 separations are presented in Fig. 2 (filled circles). The values are expressed as percent difference between test formant and background F_0 's. The mean thresholds for individual subjects lie between 0.03% and 0.89%. Thresholds for the lower CF formant were higher than those for the higher formant as in experiment 1. This difference was highly significant in a paired t -test on mean threshold estimates for the five subjects [$t(4)=3.3$, $p<0.05$].

3. Discussion

The formant matching thresholds obtained in this experiment are much smaller than the 2%–8% ΔF_0 found by other studies to be necessary to give improved identification of embedded vowels (Chalikia and Bregman, 1989; Culling and Darwin, 1993; Cutting, 1976; de Cheveigné *et al.*, 1995; Gardner and Darwin, 1986; Lea, 1992; Scheffers, 1983; Summerfield and Assmann, 1991). This sizable difference in

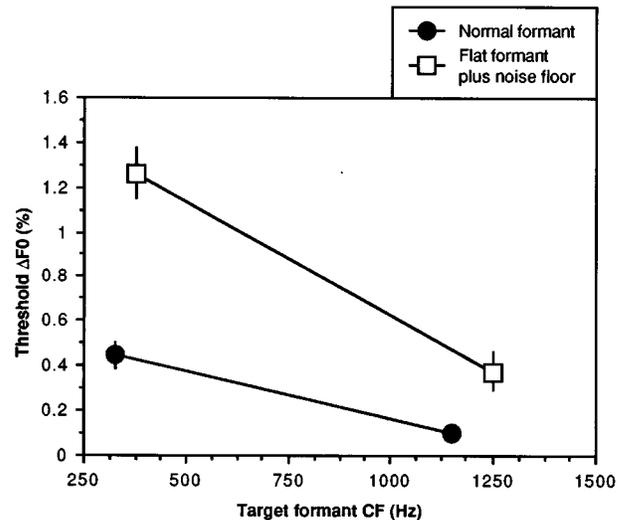


FIG. 2. Summary of polyperiodicity results for experiments 4–5 (normal formants and flat formants plus noise floor, respectively). Mean target matching thresholds, expressed as percent difference in fundamental frequency, are plotted as a function of target formant CF. Vertical bars represent \pm one standard error of the mean.

threshold may be related to the task demands. In the studies cited, subjects were required to identify vowels. In our experiment, with presentation of an isolated target formant immediately prior to the test stimulus, the decision concerning which test stimulus contained a formant matching the target would have to have been based on (most likely timbral) cues that signaled the frequency region of the test formant. These cues were quite likely derived from beats created in the region of the test components by interactions with background components. It should be noted that in the absence of the background, this task is trivial. However, even unusually small mistunings can apparently create sufficient fine-grained temporal interactions to perform the task, especially when the component spacing is small on the basilar membrane as is the case with the higher CF formant. This effect may result from interactions similar to those described for mistuned harmonics by Moore *et al.* (1985) and Hartmann *et al.* (1990).

It is interesting to compare the thresholds for this experiment with those obtained in experiment 1. Assimilating rms modulation thresholds measured in experiment 1 and the ΔF_0 thresholds from experiment 2 to the same dependent variable, a repeated measures ANOVA was performed on the mean data across repetitions for the five subjects that completed both experiments, with factors cue type (FM vs ΔF_0) and target formant (325 vs 1150 Hz). The effect of target CF was significant [$F(1,4)=32.2$, $p<0.005$], but neither cue type nor its interaction with target CF were significant [$F(1,4)=2.4$ and $F(1,4)=0.7$, respectively].¹

For the ΔF_0 stimuli of experiment 4, the beat rate between adjacent components of test formant and background would be equal to the local frequency difference between the interacting components. The amplitude modulation depth created by their interaction would be a function of their relative amplitudes, being greatest if their amplitudes were equal and decreasing monotonically as a function of their differ-

ence. In order to evaluate the potential contribution of these interactions to target matching performance, the frequency difference between adjacent components of test and background components was increased and pink noise was added in experiment 5.

B. Experiment 5: “Flat” formants in the presence of a pink noise masker

1. Stimuli and method

Experiment 4 on polyperiodicity was replicated using flat formants (target CFs of 375 and 1250 Hz) and a pink noise floor as in experiment 3. Stimuli were presented at a level of 35 dB/component. The five subjects participating in this experiment had participated in all the other experiments. The starting (maximum) values of ΔF_0 used in the adaptive procedure ranged from 0.2% to 8%. The step sizes ranged from 0.01% to 0.4%. For some subjects the maximum values and stepsizes were different for the two target formants.

2. Results

Mean thresholds are presented in Fig. 2 (open squares). Individual formant matching thresholds lie between 0.10% and 2.98% F_0 separation. Even when potential beating was greatly reduced by separating test and background components in frequency and a noise floor was added, subjects were still able to match the embedded test sounds to the target formant, though matching thresholds were quite variable across subjects and target formant CFs.

An unpaired t test on five threshold estimates for each subject (removing outlying data for one subject at the higher CF target²) showed that the difference between mean thresholds for the two formants was highly significant [$t(43)=6.12$, $p<0.0001$].

All subjects' thresholds obtained in this experiment were higher than those from experiment 4 using normal formants and no masking noise. An analysis of variance on factors stimulus type (2), target formant (2) and repetitions (5) (with outlying data removed for one subject as before) confirmed that the ΔF_0 thresholds were lower for normal formants than for flat formants in noise [$F(1,75)=46.02$; $p<0.0001$]. However, the interaction between stimulus type and target formant CF was also significant [$F(1,75)=11.28$, $p<0.005$], indicating that the decrease in threshold that accompanied an increase in target CF was greater for the flat formant stimuli with the noise floor than for the normal formants in quiet. Further, the greater reduction of beating between nearby test and background components for the lower CF target was probably responsible for the larger threshold increase at that CF (0.8% increase) than at the higher CF (0.3% increase).

3. Discussion

Though it seems to help significantly, the detection of beats produced by interacting test and background components does not appear to be entirely responsible for target matching based on polyperiodicity. This conclusion can be drawn from the fact that even after reduction of low-rate beats by limiting considerably the interactions between proximal formant and background components and by per-

turbing interactions between more distant, partially resolved components with a noise floor, subjects matched embedded formants at relatively small F_0 separations (<1.5% on average) and rms FM widths (<1% on average).

To compare the results for FM and ΔF_0 stimuli, the mean thresholds across repetitions for the two target formants and five subjects common to experiments 1, 3, 4, and 5 were submitted to a three-way analysis of variance (with missing data in experiment 3 and outlying data removed in experiment 5 as before). Thresholds for FM stimuli were expressed as rms modulation width. The independent variables were cue type (FM vs ΔF_0), stimulus type (normal formant in quiet vs flat formant in pink noise), and target formant CF (325/375 vs 1150/1250). The main effect of cue type was not significant [$F(1,31)=0.18$]. The main effect of stimulus type was significant [$F(1,31)=11.31$, $p<0.005$], reflecting the fact that the thresholds for normal formants in quiet are lower than those for flat formants presented in noise across both cue types. From comparisons of differences between thresholds for the two stimulus types, it appears that the use of flat formants against a pink-noise floor impairs target matching performance more for statically mistuned test formants than for test formants with frequency modulation at the low CF target [$t(4)=3.6$, $p<0.05$]. At the higher CF target there is no difference between FM and ΔF_0 stimuli [$t(3)=0.7$, n.s., missing data for one subject].

The larger effect of changing from normal formants in quiet to flat formants in pink noise for ΔF_0 stimuli compared to FM stimuli at the lower CF suggests a difference in the form of beating created by the two classes of cues. To examine the nature of beating present in the stimuli for the first five experiments, and to estimate the relative amount present under the different experimental conditions, we performed a computer simulation of auditory filtering and devised a measure of within-channel fluctuation of activity.

IV. COMPUTER SIMULATION OF AUDITORY BEATS

The gammatone filter model of peripheral auditory analysis (Patterson and Holdsworth, 1990; Patterson *et al.*, 1992) was used for the simulations. Synthesized stimuli from experiments 1–5 at or slightly above mean threshold were analyzed. The analysis consisted of centering gammatone filters on and between components (on an ERB-rate scale, Glasberg and Moore, 1990) in the frequency region around the test formant, and then extracting the envelope of the resulting waveform at the output of each filter by taking the Hilbert transform. As such the temporal resolution of the envelope was determined by the gammatone filter bandwidth. Example results of this process are shown in Fig. 3 for an embedded, 1150-Hz CF, “normal” formant at threshold frequency modulation width from experiment 1 and in Fig. 4 for the corresponding “flat” formant stimulus at threshold from experiment 2.

To quantify the amount of modulation present in a given auditory channel, the rms modulation index (m_{rms}) was computed as the rms amplitude of the envelope divided by its mean taken over the steady-state portion of the sound between 200 and 800 ms. This index gives a mean-independent measure of variation about the envelope mean that is com-

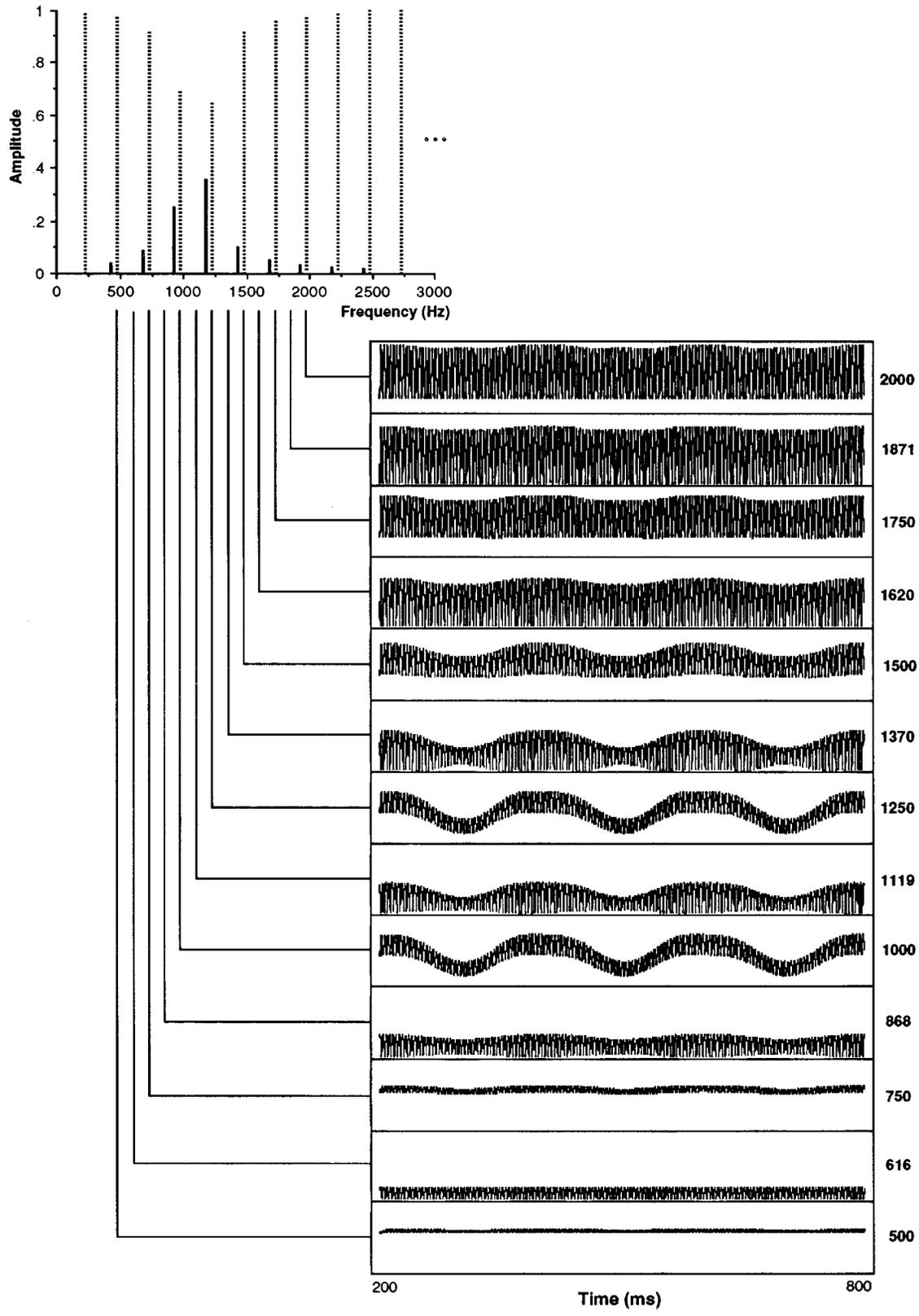


FIG. 3. Example of envelopes of the output from gammatone filtering of a stimulus in experiment 1: a frequency modulated test sound with a normal formant centered on 1150 Hz and an rms modulation width giving threshold matching (0.7%). The waveforms represent the envelope of the auditory filter output from 200 to 800 ms in the 1-s stimulus. The auditory filter CFs are listed to the right and their placement with respect to the stimulus components is indicated. Test components (solid) have been displaced with respect to background components (dotted) for visibility.

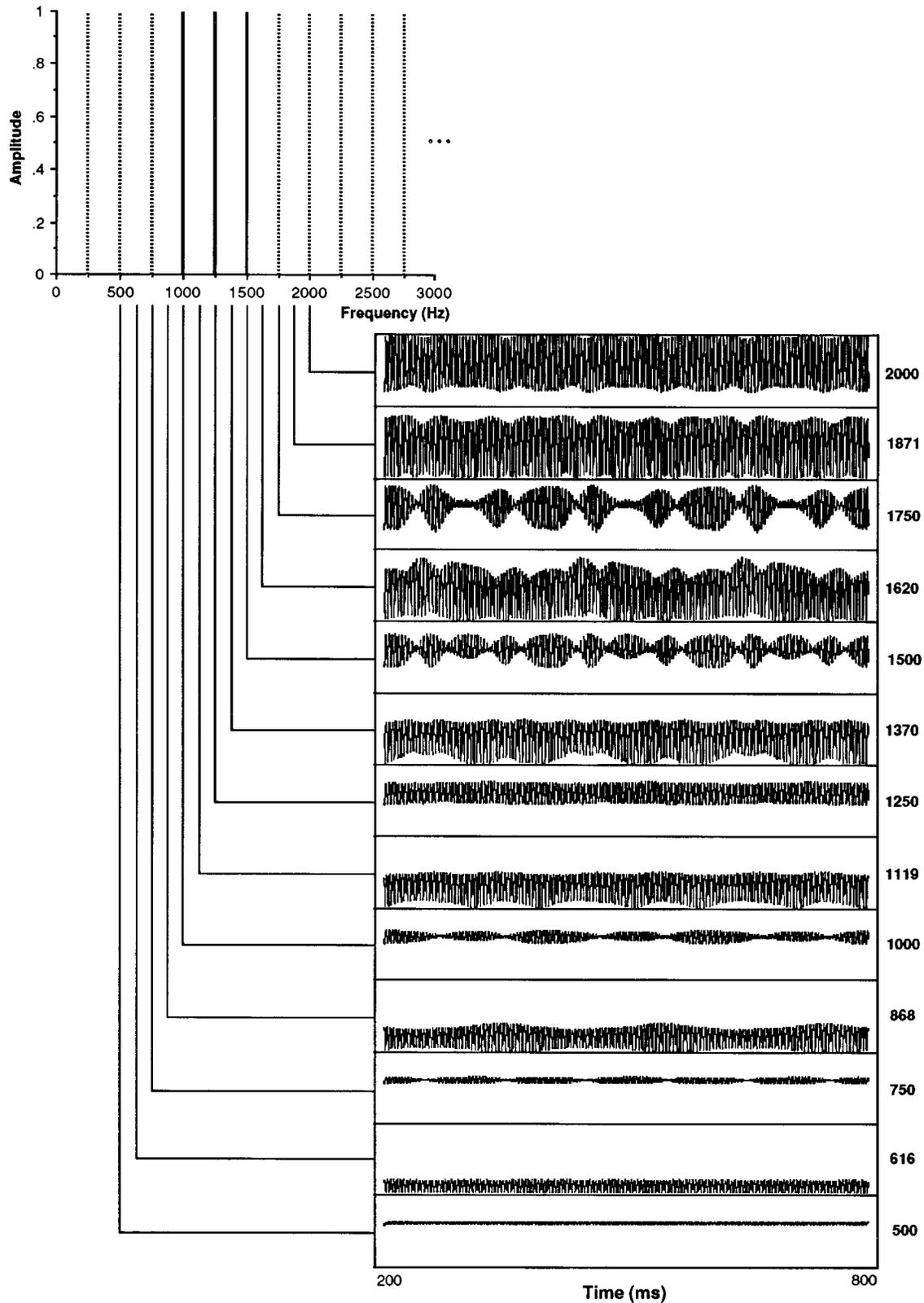


FIG. 4. Example of envelopes of the output from gammatone filtering of a stimulus in experiment 2: a frequency modulated test sound with a flat formant centered on 1250 Hz and an rms modulation width giving threshold matching (1.1%). (See Fig. 5 caption.)

parable across different, arbitrarily complex modulation functions. To represent a sort of fluctuation profile, m_{rms} was plotted as a function of filter CF on an ERB-rate scale. These profiles are shown in Fig. 5(a) for the stimuli represented in

Figs. 3 and 4. Note that m_{rms} is greatest in auditory filters centered between stimulus components. As can be seen in those figures, the preponderance of the fundamental frequency (250 Hz) in the envelope increases with increasing

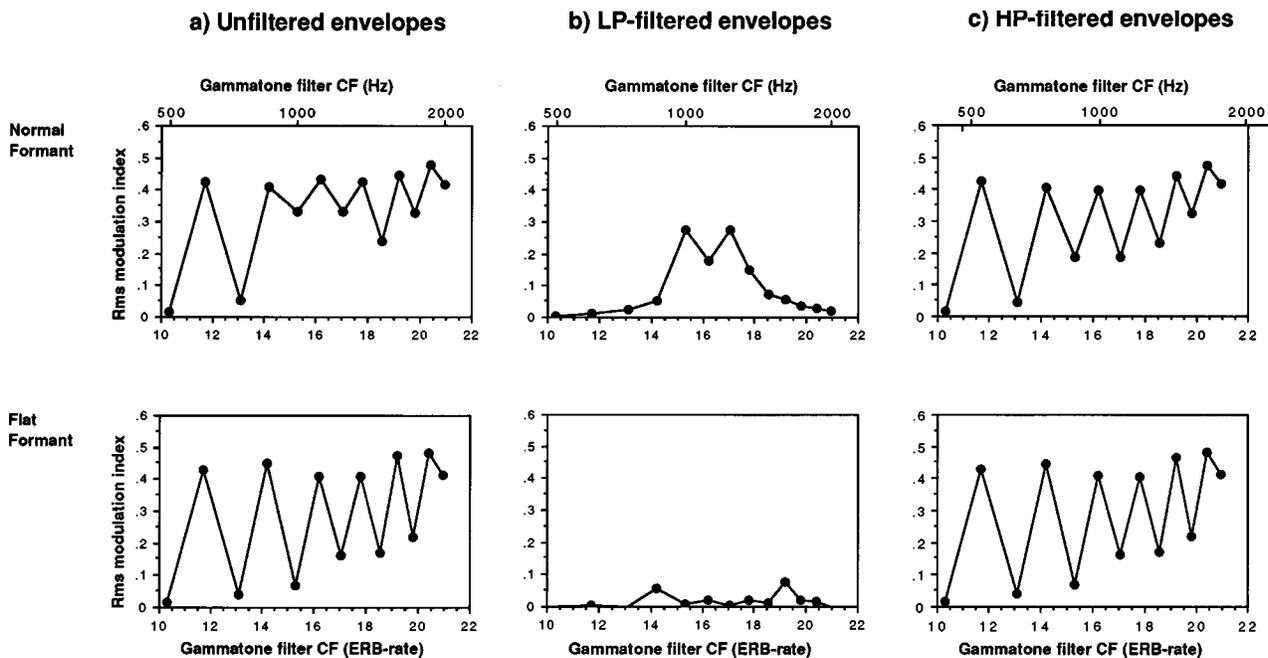


FIG. 5. Fluctuation profiles (rms modulation index, m_{rms} , as a function of auditory filter center frequency, ERB rate) for the stimuli in Figs. 3 and 4 derived from envelopes extracted from gammatone auditory filter output (a) and from the same envelopes that have been low-pass filtered at 100 Hz (b) and high-pass filtered at 150 Hz (c).

filter CF. A corresponding increase in m_{rms} is apparent in Fig. 5(a), particularly for auditory filters centered on stimulus components.

Of greatest interest for our concerns are the envelope fluctuations in the region of the formant skirts for normal formant stimuli or in the region between adjacent test and background components in the flat format stimuli. These fluctuations may result from several possible sources: (1) beating between nearby test and background components for normal formant stimuli with ΔF_0 or FM (Fig. 3), (2) induced AM as components move along the slopes of the formant skirts for normal formant stimuli with FM (Fig. 3), (3) induced AM in all FM stimuli as the modulated components move along the skirts of the auditory filters (Figs. 3 and 4), (4) beats of mistuned harmonic relations in regions between adjacent test and background components for flat formant stimuli (Fig. 4), and (5) random envelope fluctuations due to the noise floor. Of these five potential sources, the one most likely to be affected by our choice of sine phase relations among stimulus components is the fourth. The choice of random phases or other low peak-factor phase relations would decrease the modulation depth due to this cue.

To examine more closely the nature of beats created in our stimuli, fluctuation profiles were computed from the original envelope signals as well as from low-pass and high-pass versions of the envelopes. Low-rate components include sources 1–3 mentioned above, while source 4 is carried by higher frequencies in the vicinity of F_0 . Source 5 is a broader-band fluctuation source. A Kaiser filter with a 255-point impulse response and a stopband attenuation of 70 dB was applied to the extracted envelopes. A 100-Hz cutoff was used for the low-pass versions and a 150-Hz cutoff was used for the high-pass versions. The rms modulation index was then computed on these filtered envelopes. Since high-pass

filtering removes the dc component, the envelope mean used for calculating m_{rms} was taken from the unfiltered version. The result of the low-pass filtering process is shown in Figs. 6 and 7 for the same stimuli in Figs. 3 and 4.

Of the five sources of low-rate beating mentioned above, this procedure preserves all but the fourth. Compare envelopes for the 1620-Hz channel between Figs. 4 and 7. In this channel, the beating in the unfiltered version is carried by higher periodicities interacting in the envelope that are attenuated by the filtering. To the contrary, beating due to nearby components in the 1250-Hz channel in Fig. 3 is preserved in Fig. 6. Note further that the low-rate fluctuations induced by FM in the skirts of auditory filters are very small at threshold FM widths (1119- and 1370-Hz channels in Fig. 7). The effects of filtering on the fluctuation profiles are shown in Fig. 5(b) and (c). Note particularly that the modulation depth profile of lower-rate fluctuations signals the region of the normal formant but not the flat formant. The higher-rate fluctuations are prominent in both stimulus types. In general however, differences between the patterns for normal formants are most visible in profiles derived from low-pass filtered envelopes.

Results of the preceding experiments and model analyses suggest that, in some cases at least, beating in specific frequency regions may be used to perform the matching task. In these cases, low-rate sources of fluctuation seem to make the greatest distinction between test formants. Therefore, experiment 6 sought to determine the amplitude modulation index necessary to just match the spectral region of auditory beats with a low-rate fluctuation pattern similar to those revealed by the above analyses, when this was the only cue available to perform the target matching task.

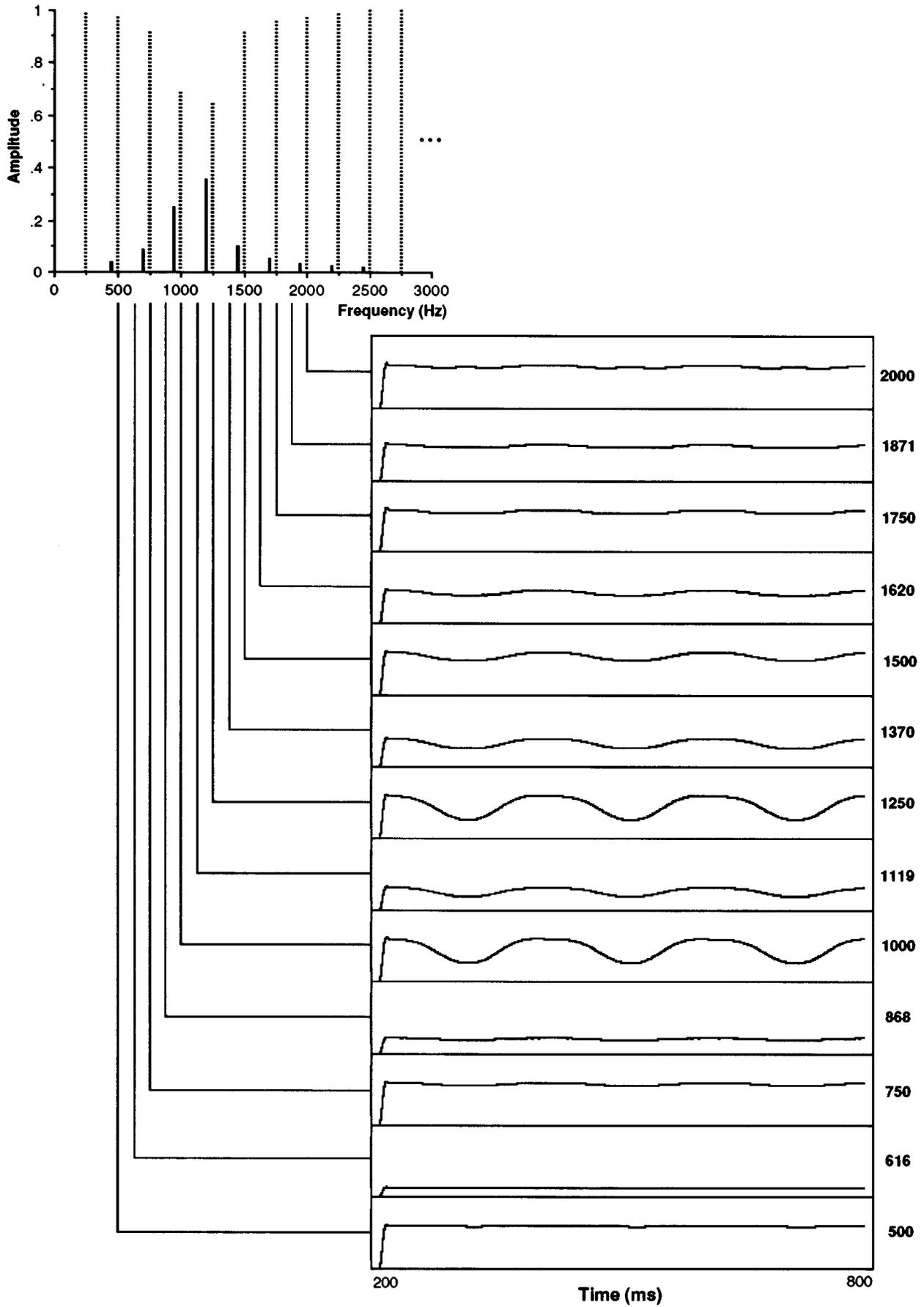


FIG. 6. 100 Hz low-pass filtered envelopes for the stimulus represented in Fig. 3.

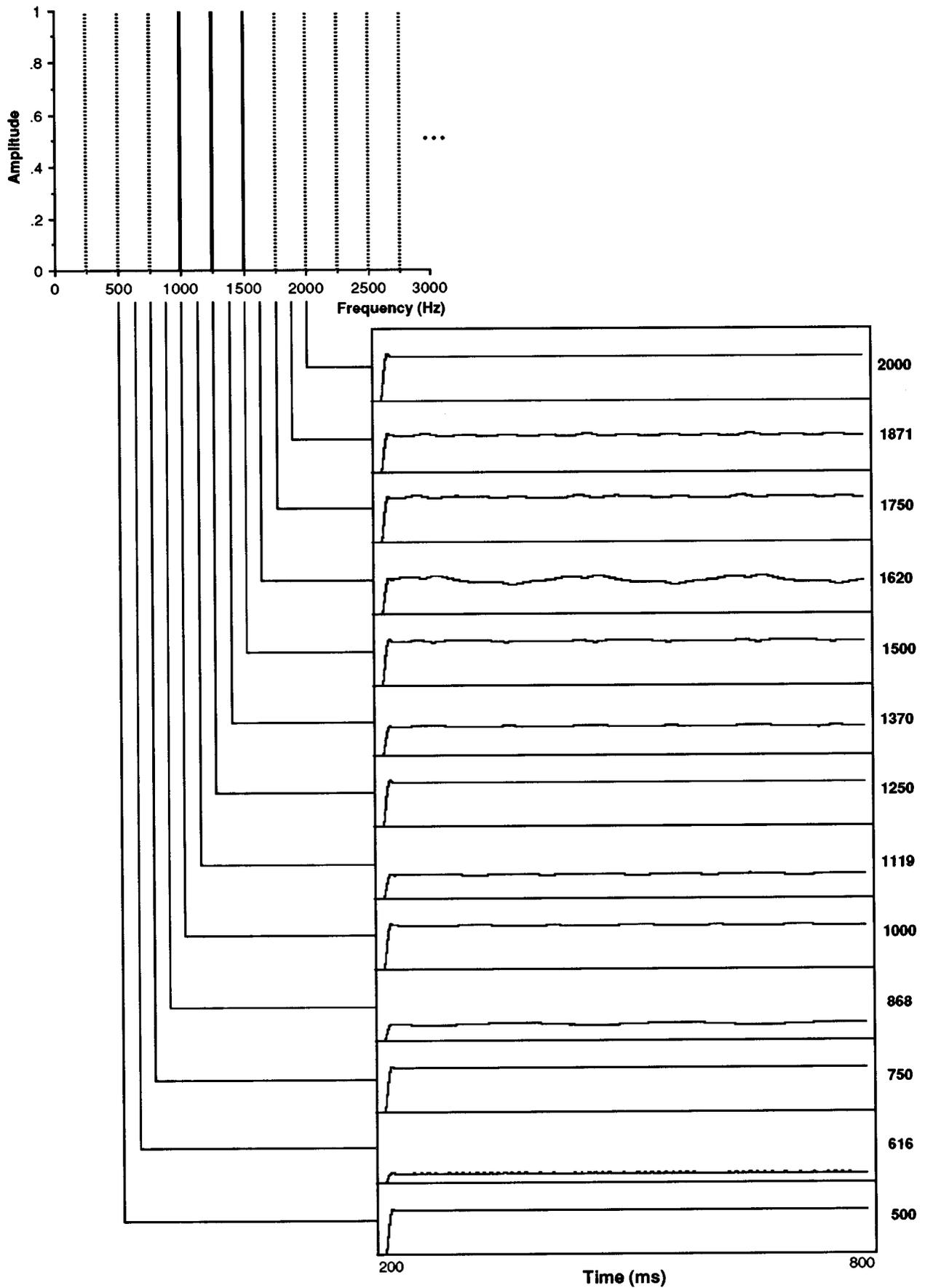


FIG. 7. 100 Hz low-pass filtered envelopes for the stimulus represented in Fig. 4.

V. DETECTION OF THE SPECTRAL REGION OF AUDITORY BEATS

A. Experiment 6: Amplitude modulation of “flat” formants

1. Stimuli and method

“Flat” formant stimuli were used in this study. The amplitudes of the background components remained fixed. The formant components were amplitude modulated with a periodic waveform designed to be representative of the low-rate beat patterns derived from simulated auditory filtering of stimuli in the previous experiments. The modulation waveform was composed of the first 13 harmonics of 5 Hz whose amplitudes corresponded to a $1/f$ spectral envelope. This spectral composition roughly reflects that found in several of the FM stimuli in the auditory simulations described in Sec. V. Fluctuation rates for ΔF_0 stimuli depend on harmonic rank but have regions with the rates included in this modulation waveform for all formant comparisons at threshold except the highest CF formant in Experiment 4. The phase relations were arbitrarily chosen from those found in the amplitude envelope of a frequency-modulated test formant with a 1700-Hz CF after passing through an auditory filter centered on 1500 Hz. The AM was applied such that the mean level for each component of the test formant over the duration of the tone corresponded to the fixed level of the background components. This ensured that all test stimuli had equal rms levels. For these stimuli, low-rate AM is localized only in the region of the flat formant. The 250-Hz fluctuations are still present in channels between lower harmonic components and in all higher-frequency channels.

A 16-turn adaptive tracking procedure was used with interleaved presentation of the three target CFs (375, 750, 1250 Hz). The dependent variable was the rms modulation index (m_{rms}). This value varied from 0 to 0.3 in steps of 0.03. After training, five threshold estimates for each target formant were collected for two highly trained subjects (the authors).

2. Results

The results show that subjects could identify the spectral region in which beating was located and use this cue to perform the target matching task. The rms modulation index needed for this identification was in the range 0.11–0.22. The means were 0.15, 0.16, 0.15 for target formants of 375, 750, and 1250 Hz, respectively. Tukey–Kramer comparisons across the five threshold estimates for each subject showed that the differences in mean threshold between target CFs were not significant (the critical difference at $p=0.05$ was 0.075).

3. Discussion

Thresholds for detection of beating between two closely spaced frequency components (roughly equivalent to sine wave AM) have been measured by Riesz (1928) and Viemeister (1979). For a carrier frequency of 1 kHz and a level of 50 dB SPL, the modulation index at detection threshold varies from 0.03 to 0.12 (0.02–0.08 in terms of m_{rms} presuming sinusoidal modulation) depending on the modulation fre-

quency (summary data estimated from Viemeister, 1979, Fig. 1). The thresholds for formant matching in the present experiment are about a factor of 3 higher than Riesz’s and Viemeister’s beat detection thresholds. This comparison suggests that modulation detection was not sufficient in itself to result in matching of the frequency region where the modulation occurred.

More recently, Hall and Grose (1991) and Moore and Bacon (1993) have investigated subjects’ abilities to identify which frequency component in a complex sound (two or six harmonic components) was sinusoidally amplitude modulated. Hall and Grose presented two-component signals with frequencies of 1 and 2 kHz and asked subjects to identify which component was amplitude modulated. Thresholds gave m_{rms} values of 0.05 to 0.27. Moore and Bacon found that subjects’ identification performance was well above chance with m_{rms} values of 0.35 and 0.71, when a probe component was presented either before or after the complex signal containing a single modulated harmonic. The values from both of these studies agree with our thresholds. In addition, Moore and Bacon tested identification of single, modulated components in six-component complexes. Threshold m_{rms} values varied from 0.14 to 0.22, indicating that sensitivity may be greater to the “odd man out” in multicomponent backgrounds than in the presence of a single unmodulated component. The authors suggest that the modulation of a single component may promote its perceptual segregation, but that it is more difficult to then decide which component was modulated in the two-component complex than in the six-component complex.

B. Comparison of model output across experiments

Of interest for the present study is the amount of modulation in the region of the test formant that exceeds that needed for threshold matching when such modulation is the only source of information. In fluctuation profiles for threshold stimuli in experiment 6, the 0.15 m_{rms} value was recovered perfectly in the low-pass filtered envelopes in channels centered on or between the modulated components. Such was not the case in the unfiltered envelopes in which m_{rms} was contaminated with the fundamental frequency or in the high-pass filtered envelopes in which only the components near 250 Hz were visible. In order to be able to compare these simulations with those for stimuli presented in noise, a pink noise floor was also added to threshold AM stimuli and the fluctuation profile was computed from low-pass and high-pass filtered envelopes.³

Figures 8 and 9 present the m_{rms} values computed on low-pass and high-pass filtered envelopes, respectively, at each CF for a given test formant in each experiment after the corresponding values derived from threshold stimuli in experiment 6 have been subtracted (Δm_{rms}). Thus, values below zero are interpreted as being too small to allow formant matching on the basis of this cue alone. If Δm_{rms} is greater than zero, this cue may be useful if the fluctuation profiles for the two formants are sufficiently different. For stimuli from experiments 1, 2, and 4 (quiet) the value 0.15 was simply subtracted at all CFs since this threshold was nearly constant across test formants. Profiles from the stimuli of

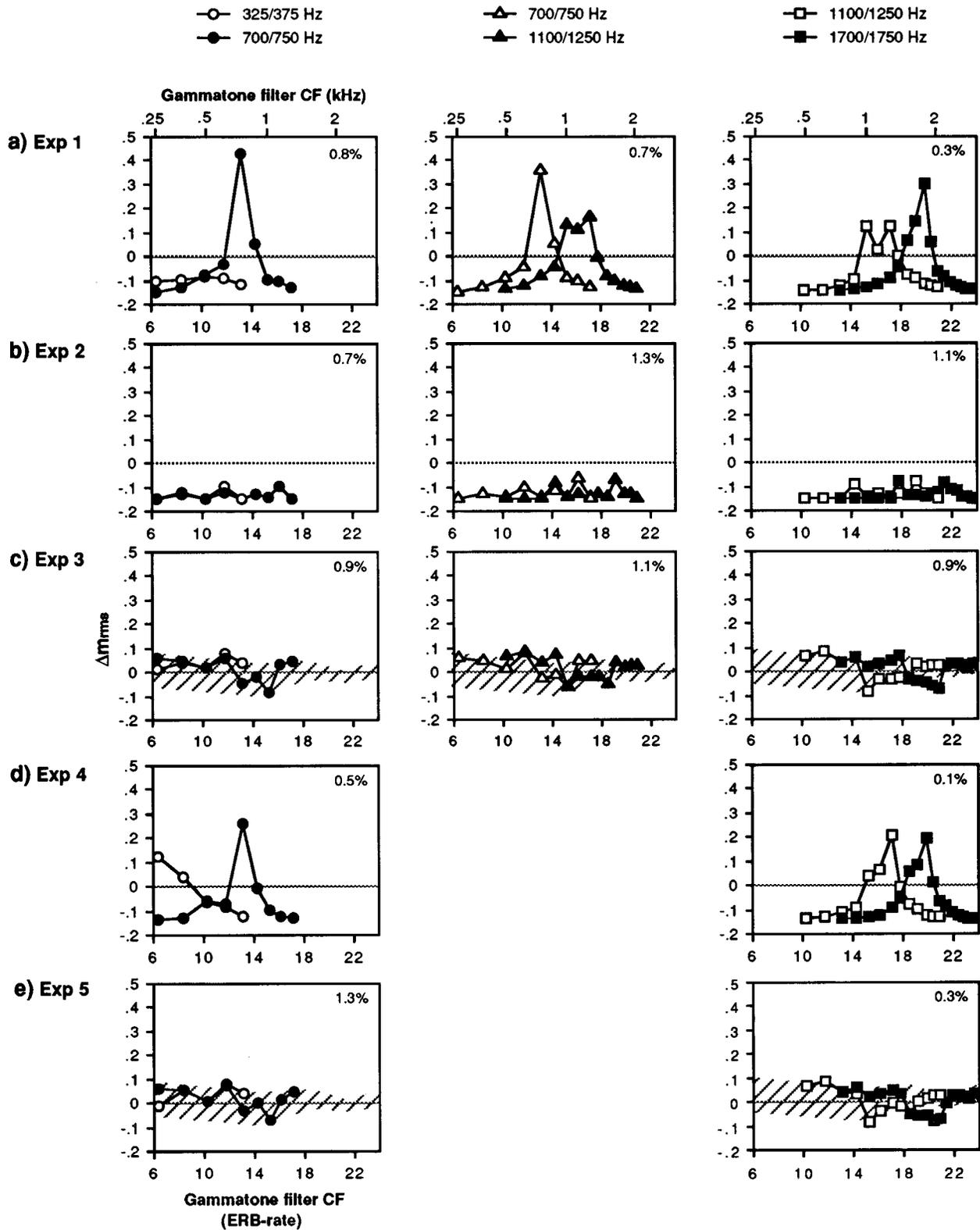


FIG. 8. Differences in rms modulation index (Δm_{rms}) between threshold stimuli in experiments 1–5 and those in experiment 6 (with or without a pink noise floor as appropriate) as measured from model output. The difference is plotted as a function of the CF of the gammatone filter. Pairs of test formants are presented in columns and experiments are presented in rows (a)–(e). The m_{rms} values (see text) are computed on low-pass filtered envelopes for gammatone filters centered on or between harmonics in the frequency region around the test formant. The textured region indicates the range of variation in m_{rms} across the ten different pink noise samples used in experiments 3 and 5. The values in the upper right-hand corner of each panel are the threshold rms FM width or ΔF_0 used in the analysis.

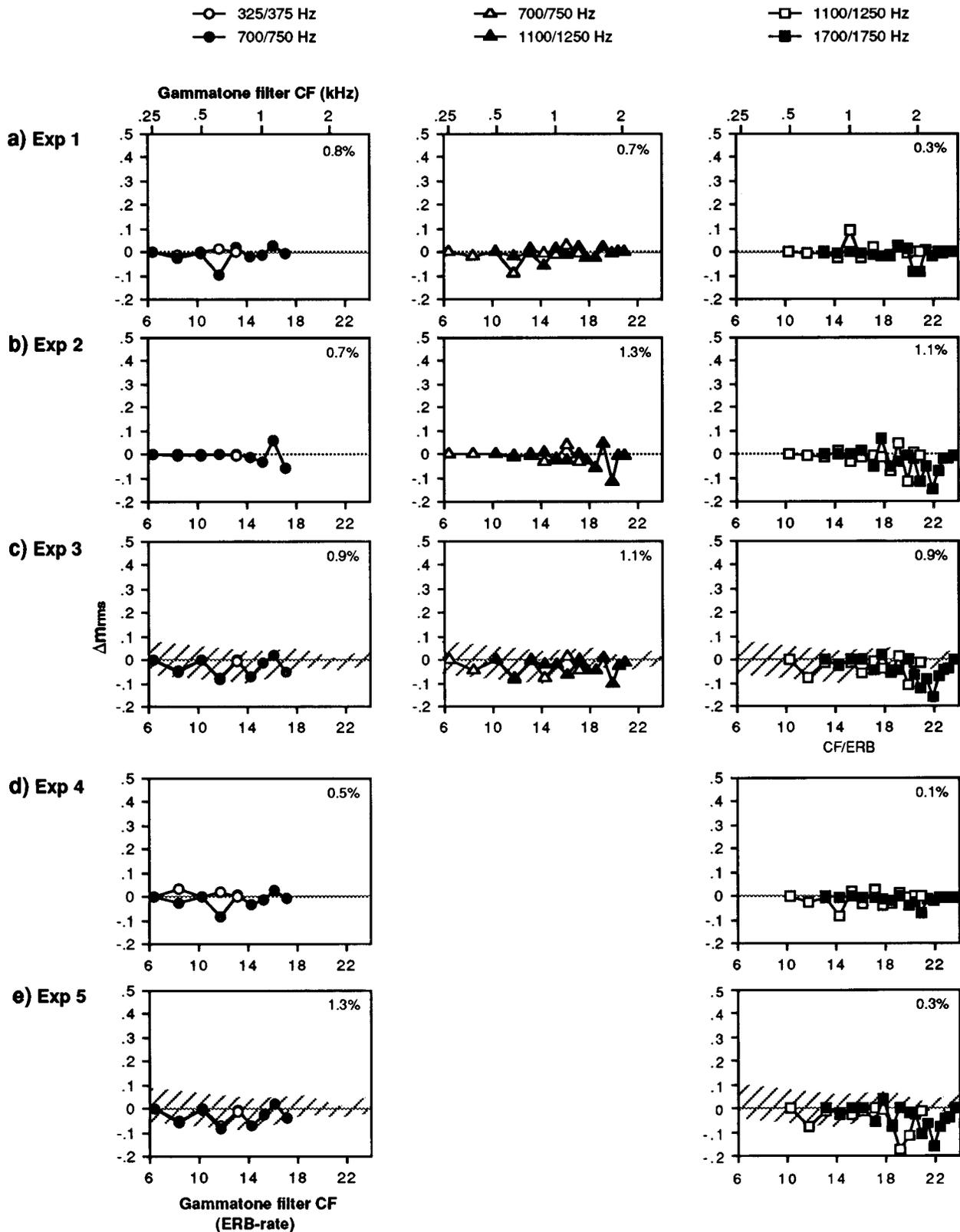


FIG. 9. Δm_{rms} for threshold stimuli in experiments 1–5 computed on high-pass filtered envelopes (see Fig. 8 caption).

experiments 3 and 5 (pink noise floor) were derived by subtracting the corresponding profiles for each threshold test stimulus in experiment 6 presented against a noise floor. This latter computation is not totally adequate since the profile would vary in detail on a stimulus-to-stimulus basis as the

noise floor waveform varied. Recall that ten separate noise samples were used and were randomly selected for each stimulus presentation. In order to give a sense of the possible effect of the different noise samples, the range of variation in Δm_{rms} across the ten noise waveforms is indi-

cated as the textured region in Fig. 8(c) (experiment 3) and (e) (experiment 5). For our interpretations below, we consider points lying within this noise range to be unreliable for performing the task since they could be confounded with those obtained from the noise floor in the absence of any harmonic stimulus. Further, this representation probably overestimates sensitivity to modulation depth for higher-rate fluctuations which Viemeister (1979) has shown to be less than for low-rate modulation.

It is of particular interest to compare the figures for FM or ΔF_0 across conditions expected to reduce beating [Fig. 8(a)–(c) and (d)–(e), respectively]. For normal-formant FM stimuli at threshold (experiment 1), one or both formants create a significant amount of AM in the channels centered near the test and background components that are nearly equal in amplitude. The channels in which this AM occurs are different for each formant and discrimination of these fluctuation patterns is a likely cue for target matching. A similar explanation seems tenable for normal-formant ΔF_0 stimuli (experiment 4). Note that for threshold performance for both ΔF_0 and FM, Δm_{rms} is largely above zero, being greater than 0.2 for at least one of two formants in all cases. This suggests that the effectiveness of a pure modulation index-related cue is less in these stimuli than in simple AM stimuli.

Now we can contrast these curves with those representing the envelope fluctuations present in stimuli designed to reduce low-rate beating (experiments 2, 3, and 5). Note that the Δm_{rms} values are well below zero for all formants in experiment 2. This result would suggest that the beating that is present is not useful for target matching. For stimuli with noise floors, the Δm_{rms} values vary around zero but rarely exceed the range of variation due to the different noise samples. While this might suggest some residual detectability, we feel we are justified in rejecting this hypothesis for three reasons. First, the actual pattern of variation differs from one noise sample to the next and the variation between samples is approximately equivalent to that found between the two formants being compared. As such, it is difficult to imagine that the difference in fluctuation profile shown in Fig. 8 would allow reliable matching across noise samples. Second, for a given noise sample, the m_{rms} values do not increase monotonically with the values of either dependent variable (rms modulation width or ΔF_0), indicating that there is no reliable information in the depth of low-rate fluctuation that would allow the subject to perform the task. Thirdly, the differences in profile between pairs of test formants at threshold values of the dependent variable are very small compared to those in threshold stimuli from experiments 1 and 4 [compare (a) with (c) and (d) with (e) in Fig. 8]. It thus seems unlikely that detection of low-rate fluctuations was used to perform the matching task for flat formant stimuli.

Nonetheless, closer examination of the unfiltered model outputs resulting from flat-formant and flat-formant-plus-pink-noise stimuli reveals some apparent periodic fluctuations in channels at and above the one centered on the highest component of the test formant. In this region, there appears to be an interaction between a 250-Hz periodicity resulting from overlapping background components and ei-

ther a slightly faster periodicity resulting from overlapping test components in the ΔF_0 stimuli, or a separate periodicity that modulates around this value resulting from overlapping test components in the FM stimuli. Similar interactions between envelope periodicities are found on the lower frequency side of higher-CF, flat-formant test sounds. In this way, a lower-rate envelope fluctuation results from interaction of two higher-rate envelope fluctuations, a phenomenon perhaps akin to beats of mistuned consonances (Plomp, 1976). Comparing fluctuation profiles for high-pass filtered envelopes (Fig. 9), it becomes clear that sufficient Δm_{rms} values are obtained for threshold stimuli only in experiments 1, 2, and 4. However, the differences between profiles for the pair of test formants are very small in all cases. In experiments 3 and 5 (pink noise floors), the fluctuation profiles do not exceed the variation due to the noise floor. So detection of modulation due to higher-rate fluctuations would not appear to be a useful cue for formant matching.

We cannot completely rule out, on the basis of these data, the possibility that the detection of differences in periodicity and shape of the higher-rate envelope fluctuations may have allowed detection of the frequency region of the test formant and thus its matching to the target formant in cases in which the modulation index itself did not distinguish between test formants. However, results from experiment 6 suggest that detection or discrimination of such differences may not be sufficient in the present target matching paradigm. Further, work by Viemeister (1979) on temporal modulation transfer functions for broadband noise suggests that sensitivity to modulation frequencies around 250 Hz is much less than that at low modulation frequencies. It seems likely that greater differences would be needed to perform this task.

VI. CONCLUSIONS

- (1) Thresholds for matching spectrally embedded, frequency modulated targets are higher than FM detection thresholds for nonembedded sounds, suggesting that detectability of FM is not sufficient to perform this task.
- (2) Elimination of low-rate beats due to both interactions of nearby test and background components and tracing of formant spectral envelopes increased matching thresholds for higher-CF target formants that were modulated in frequency. Introduction of a noise floor gave no further increase in thresholds. Beating induced by frequency modulation of target components would thus seem to contribute significantly to target matching performance.
- (3) Target matching thresholds with fundamental frequency mistuning of normal formants are much lower than those found for vowel identification, most likely because a simple detection of the frequency region of beating between test formant and background would allow performance of this task. The particularly low threshold for higher CF formants (0.1% mistuning) suggests that subjects used fine-grained temporal information induced by the mistuning, although computer simulations suggest that the modulation depth itself is insufficient to explain this effect.

- (4) Elimination of low-rate beating and the addition of a noise floor significantly increased ΔF_0 thresholds and all the more so for lower CF targets. This increase at the lower CF target was much greater for ΔF_0 stimuli than for FM stimuli, suggesting that modulated targets create additional, low-rate fluctuation cues not available with static targets in stimuli with partially resolved harmonics. It is likely that this cue is related to the AM induced in auditory filters by the FM.
- (5) Comparisons of experimental measures of threshold AM depth necessary to perform the matching task with measures of within-channel modulation index derived from computer simulations of auditory filtering suggested that low-rate beats contributed to performance with normal formants but were unlikely to have played a role for flat formant stimuli.
- (6) Therefore, low-rate beating in specific frequency regions is a likely cue for the perception of frequency modulated harmonic target sources that are embedded in an interfering harmonic background and whose spectral characteristics are known or predictable. One wonders whether this mechanism may not be responsible for the increased perceptual prominence found for frequency modulated vowels in multivowel mixtures that cannot be explained on the basis of segregation (Culling and Summerfield, 1995; McAdams, 1989). Further, low-rate beats are a likely cue for detecting embedded harmonic targets that are mistuned with respect to the background. However, in the absence of this cue, target sounds can still be perceived at sufficient mistuning from the background, suggesting additional contributions of either the pattern of higher-periodicity beating (mistuned consonances) and/or perceptual segregation on the basis of polyperiodicity itself.

ACKNOWLEDGMENTS

Bennett Smith provided invaluable help in programming the experimental routines. John Holdsworth introduced us to the trials and tribulations of auditory modeling. Chris Darwin and Jean-Sylvain Liénard offered critical advice at an early stage of the experimentation. Laurent Demany, John Grose, Brian Moore, Alain de Cheveigné, and an anonymous reviewer provided helpful criticisms of early versions of the manuscript. C. M. H. M. was supported by a doctoral fellowship from the D. C. A. N. of the French Ministry of Defense. S. M. was supported in part by a grant from the Cognitive Sciences program of the French Ministry of Research and Space. Some of the experiments reported here were conducted in partial fulfillment of the requirements for C. M. H. Marin's (1991) doctoral dissertation which is available in French. Correspondence and requests for reprints should be addressed to either author at IRCAM; e-mail: marin@ircam.fr, smc@ircam.fr

¹If experiment 1 thresholds are expressed as maximum percent mistuning attained by the FM waveform, the effect of cue type just misses attaining significance [$F(1,4)=6.9, p=0.058$]. No other differences occur between the two analyses.

²For the 1250-Hz target in experiment 5, the mean threshold for one subject

was a factor of 8 (6.7 s.d.) higher than the mean of the other four subjects. All other subject means were within 1.5 s.d. of the group mean. As such we considered this data point an outlier and removed it from subsequent data analyses.

³It is, of course, possible that the thresholds would not be identical in quiet and in noise. However, the spectra of the filtered envelopes for the AM stimuli in noise still reveal prominent components at the multiples of 5 Hz in channels centered on the modulated components.

- Bregman, A. S. and Doehring, P. (1984). "Fusion of simultaneous tonal glides: The role of parallelness and simple frequency relations," *Percept. Psychophys.* **36**, 251–256.
- Carlyon, R. P. (1991). "Discriminating between coherent and incoherent frequency modulation of complex tones," *J. Acoust. Soc. Am.* **89**, 329–340.
- Carlyon, R. P. (1992). "The psychophysics of concurrent sound segregation," *Philos. Trans. R. Soc. London, Ser. B* **336**, 347–355.
- Chalikia, M. H., and Bregman, A. S. (1989). "The perceptual segregation of simultaneous auditory signals: Pulse train segregation and vowel segregation," *Percept. Psychophys.* **46**, 487–496.
- Culling, J. F., and Darwin, C. J. (1993). "Perceptual separation of simultaneous vowels: within and across-formant grouping by F_0 ," *J. Acoust. Soc. Am.* **93**, 3454–3467.
- Culling, J. F., and Summerfield, Q. (1995). "The role of frequency modulation in the perceptual segregation of concurrent vowels," *J. Acoust. Soc. Am.* **98**, 837–846.
- Cutting, J. E. (1976). "Auditory and linguistic processes in speech perception: inferences from six fusions in dichotic listening," *Psychol. Rev.* **83**, 114–140.
- de Cheveigné, A. (1993). "Separation of concurrent harmonic sounds: Fundamental frequency estimation and a time-domain cancellation model of auditory processing," *J. Acoust. Soc. Am.* **93**, 3271–3290.
- de Cheveigné, A., McAdams, S., Laroche, J., and Rosenberg, M. (1995). "Identification of concurrent harmonic and inharmonic vowels: A test of the theory of harmonic cancellation and enhancement," *J. Acoust. Soc. Am.* **97**, 3736–3748.
- Demany, L., and Semal, C. (1989). "Detection thresholds for sinusoidal frequency modulation," *J. Acoust. Soc. Am.* **85**, 1295–1301.
- Gardner, R. B., and Darwin, C. J. (1986). "Grouping of vowel harmonics by frequency modulation: Absence of effects on phonemic categorization," *Percept. Psychophys.* **40**, 183–187.
- Glasberg, G. R., and Moore, B. C. J. (1990). "Derivation of auditory filter shapes from notched-noise data," *Hear. Res.* **47**, 103–138.
- Goldstein, J. L. (1967). "Auditory nonlinearity," *J. Acoust. Soc. Am.* **41**, 676–689.
- Hall, J. W., and Grose, J. H. (1991). "Some effects of auditory grouping factors on modulation detection interference (MDI)," *J. Acoust. Soc. Am.* **90**, 3028–3035.
- Hartmann, W. M., and Klein, M. A. (1981). "The effect of uncertainty on the detection of frequency modulation at low modulation rates," *Percept. Psychophys.* **30**, 417–424.
- Hartmann, W. M., McAdams, S., and Smith, B. K. (1990). "Hearing a mistuned harmonic in an otherwise periodic complex tone," *J. Acoust. Soc. Am.* **88**, 1712–1724.
- Helmholtz, H. L. F. (1885). *On the Sensations of Tone as a Physiological Basis for the Theory of Music*, English transl. by A. J. Ellis from 4th German ed. (1877); (republ. 1954, Dover, New York).
- Lea, A. (1992). "Auditory models of vowel perception," Unpublished doctoral dissertation, University of Nottingham, UK.
- Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467–477.
- Marin, C. M. H. (1991). "Processus de séparation perceptive des sources sonores simultanées," Unpublished doctoral dissertation, Université de Paris III, France.
- Marin, C. M. H. (1996). "Is fundamental frequency separation or harmonic coincidence a cue contributing to concurrent sound segregation?" (submitted).

- Marin, C. M. H., and McAdams, S. (1990). "Segregation of concurrent sounds. II: Effects of spectral envelope tracing, frequency modulation coherence and frequency modulation width," *J. Acoust. Soc. Am.* **89**, 341–351.
- McAdams, S. (1984). "Spectral Fusion, Spectral Parsing, and the Formation of Auditory Images," Unpublished Ph.D. dissertation, Stanford University, Stanford, CA.
- McAdams, S. (1989). "Segregation of concurrent sounds. I: Effects of frequency modulation coherence," *J. Acoust. Soc. Am.* **86**, 2148–2159.
- McAdams, S., and Marin, C. M. H. (1990). "Auditory processing of frequency modulation coherence," in *Proceedings of the 6th Annual Meeting of the International Society of Psychophysics* (I.S.P., Würzburg), pp. 175–180.
- Moore, B. C. J., and Bacon, S. (1993). "Detection and identification of a single modulated carrier in a complex sound," *J. Acoust. Soc. Am.* **94**, 759–768.
- Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (1985). "Thresholds for the detection of inharmonicity in complex tones," *J. Acoust. Soc. Am.* **77**, 1861–1867.
- Patterson, R., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., and Allerhand, M. (1992). "Complex sounds and auditory images," in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, Oxford), pp. 429–446.
- Patterson, R. D. and Holdsworth, J. (1990). "An Introduction to Auditory Sensation Processing" (unpublished manual), MRC Applied Psychology Unit, Cambridge.
- Plomp, R. (1976). *Aspects of Tone Sensation: A Psychophysical Study* (Academic, London).
- Riesz, R. R. (1928). "Differential sensitivity of the ear for pure tones," *Phys. Rev.* **31**, 867–875.
- Scheffers, M. T. M. (1983). "Sifting Vowels: Auditory Pitch Analysis and Sound Integration," Unpublished doctoral dissertation, University of Groningen, The Netherlands.
- Smith, B. K., and Chervin, P. (1986). "Boris: An application of the Fujitsu MB 8754 DSP chip," in *Proceedings of the 1986 International Computer Music Conference, Den Haag* (Computer Music Association, San Francisco).
- Summerfield, Q., and Assmann, P. F. (1991). "Perception of concurrent vowels: effects of harmonic misalignment and pitch-period asynchrony," *J. Acoust. Soc. Am.* **89**, 1364–1377.
- Summerfield, Q., and Culling, J. F. (1992). "Auditory segregation of competing voices: absence of effects of FM or AM coherence," *Philos. Trans. R. Soc. London, Ser. B* **336**, 357–366.
- Vercoe, B. (1986). *Csound: A Manual for the Audio Processing System and Supporting Programs* (Media Lab, MIT, Cambridge, MA).
- Viemeister, N. F. (1979). "Temporal modulation transfer functions based upon modulation thresholds," *J. Acoust. Soc. Am.* **66**, 1364–1380.